

AR Spec 1 Spec 3 Spec 5 Spec 7 CoSpec

Temp: 0

Temp: 0.5

Temp: Random

Capacity (reqs/s)

P99 TPOT (s/token)

